

TITLE

PLANT AMINO ACID BIOSYNTHETIC ENZYMES

This application is a continuation-in-part of Application No. 09/424,976 filed on December 2, 1999 which is a national stage application of PCT/US98/12073 with an International filing date of June 11, 1998, which in turn claims priority benefit of U.S. Provisional Application No. 60/049406, filed June 12, 1997 and U.S. Provisional Application No. 60/065385, filed November 12, 1997.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes involved in amino acid biosynthesis in plants and seeds.

BACKGROUND OF THE INVENTION

Many vertebrates, including humans, lack the ability to manufacture a number of amino acids and therefore require these amino acids in their diet. These are called essential amino acids. Grain-derived foods or feed, however, are deficient in certain essential amino acids, such as lysine, the sulfur-containing amino acids methionine and cysteine, threonine and tryptophan. For example, in corn (*Zea mays L.*) lysine is the most limiting amino acid for the dietary requirements of many animals, and soybean (*Glycine max L.*) meal is used as an additive to corn-based animal feeds primarily as a lysine supplement. Often microbial-fermentation produced lysine is needed for such supplementation. Thus, an increase in lysine content of either corn or soybean would reduce or eliminate the need to supplement mixed grain feeds with lysine produced via fermentation.

Furthermore, in corn the sulfur amino acids are the third most limiting amino acids, after lysine and tryptophan, for the dietary requirements of many animals. Legume plants, however, while rich in lysine and tryptophan, have low sulfur-containing amino acid content. Therefore, the use of soybean meal to supplement corn in animal feed is not satisfactory. An increase in the sulfur amino acid content of either corn or soybean would improve the nutritional quality of the mixtures and reduce the need for further supplementation through addition of more expensive methionine.

One approach to increasing the nutritional quality of human foods and animal feed is to increase the production and accumulation of specific free amino acids via genetic engineering of the biosynthetic pathway of the essential amino acids. Biosynthetically, lysine, threonine, methionine, cysteine and isoleucine are all derived from aspartate. Regulation of the biosynthesis of each member of this family is interconnected (see Figure 1). The organization of the pathway leading to biosynthesis of lysine, threonine, methionine, cysteine and isoleucine indicates that over-expression or reduction of expression of genes encoding, *inter alia*, aspartic semialdehyde dehydrogenase, homoserine kinase,

diaminopimelate decarboxylase, cysteine synthase and cystathionine β -lyase in corn and soybean could be used to alter levels of these amino acids in human food and animal feed. However, few of the genes encoding enzymes that regulate this pathway in plants, especially corn and soybeans, are available. Accordingly, availability of nucleic acid sequences encoding all or a portion of these enzymes would facilitate development of nutritionally improved crop plants.

SUMMARY OF THE INVENTION

The present invention relates to isolated polynucleotides selected from the group consisting of SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, 8, 10, 12, 14, 16, 18, 53, 55, 21, 23, 25, 27, 58, 30, 61, 63, 33, 35, 37, 39, 67, 69, and 71.

The present invention concerns isolated polynucleotides comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51; (b) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:9, 11, 13, 15, 17, 19, 54 and 56; (c) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:22, 24, 26, 28, and 59; (d) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:31, 62, and 64; and (e) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72. It is preferred that the identity be at least 85%, more preferably at least 90%, still more preferably at least 95%. This invention also relates to the isolated complement of such polynucleotides, wherein the complement and the polynucleotide consist of the same number of nucleotides, and the nucleotide sequences of the complement and the polynucleotide have 100% complementarity.

In a third embodiment nucleotide sequence of the isolated first polynucleotide is selected from SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, SEQ ID NOs:8, 10, 12, 14, 16, 18, 53 and 55, SEQ ID NOs:21, 23, 25, 27, and 58, SEQ ID NOs:30, 61, and 63, and SEQ ID NOs:33, 35, 37, 39, 67, 69, and 71.

In a fourth embodiment, this invention concerns an isolated polynucleotide encoding an aspartic semialdehyde dehydrogenase, a diaminopimelate decarboxylase, a homoserine kinase, a cysteine γ synthase or a cystathionine β -lyase.

In a fifth embodiment, this invention relates to a chimeric gene comprising the polynucleotide of the present invention.

In a sixth embodiment, the present invention concerns an isolated nucleic acid molecule that comprises at least 180 nucleotides and remains hybridized with the isolated polynucleotide of the present invention under a wash condition of 0.1X SSC, 0.1% SDS, and 65°C.

In a seventh embodiment, the invention also relates to a host cell comprising a chimeric gene of the present invention or an isolated polynucleotide of the present invention. The host cell may be eukaryotic, such as a yeast cell or a plant cell, or prokaryotic, such as a bacterial cell. The present invention may also relate to a virus comprising an isolated polynucleotide of the present invention or a chimeric gene of the present invention.

In an eighth embodiment, the invention concerns a transgenic plant comprising a polynucleotide of the present invention.

In a ninth embodiment, the invention relates to a method for transforming a cell by introducing into such cell the polynucleotide of the present invention, or a method of producing a transgenic plant by transforming a plant cell with the polynucleotide of the present invention and regenerating a plant from the transformed plant cell.

In a tenth embodiment, the invention concerns a method for producing a nucleotide fragment by selecting a nucleotide sequence comprised by a polynucleotide of the present invention and synthesizing a polynucleotide fragment containing the nucleotide sequence. It is understood that the nucleotide fragment may be produced *in vitro* or *in vivo*.

In an eleventh embodiment the invention concerns an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a polypeptide of at least 60 amino acids and having a sequence identity of at least 80% based on the Clustal method of alignment when compared to an amino acid sequence selected from the group consisting of SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51; (b) a polypeptide of at least 60 amino acids having a sequence identity of at least 95% based on the Clustal method of alignment when compared to an amino acid sequence selected from the group consisting of SEQ ID NOs:9, 11, 13, 15, 17, 19, 54 and 56; (c) a polypeptide of at least 60 amino acids having a sequence identity of at least 80% based on the Clustal method of alignment when compared to an amino acid sequence selected from the group consisting of SEQ ID NOs:22, 24, 26, 28, and 59; (d) polypeptide of at least 60 amino acids having an identity of at least 95% based on the Clustal method of alignment when compared to an amino acid sequence selected from the group consisting of SEQ ID NOs:31, 62, and 64; and (e) a polypeptide of

at least 60 amino acids having a sequence identity of at least 85% based on the Clustal method of alignment when compared to an amino acid sequence selected from the group consisting of SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72. It is preferred that the identity be at least 85%, it is more preferred if the identity is at least 90%, it is preferable that the identity be at least 95%.

In a twelfth embodiment the invention relates to an isolated polypeptide selected from SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51, SEQ ID NOs:9, 11, 13, 15, 17, 19, 54 and 56, SEQ ID NOs:22, 24, 26, 28, and 59, SEQ ID NOs:31, 62, and 64, and SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72.

In a thirteenth embodiment, this invention concerns an isolated polypeptide having aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine γ synthase, or cystathionine β -lyase function.

In a fourteenth embodiment, this invention relates to a method of altering the level of expression of a plant biosynthetic enzyme in a host cell comprising: transforming a host cell with a chimeric gene of the present invention; and growing the transformed host cell under conditions that are suitable for expression of the chimeric gene.

A further embodiment of the instant invention is a method for evaluating a compound for its ability to inhibit the activity of a plant biosynthetic enzyme selected from the group consisting of aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine γ synthase and cystathionine β -lyase, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a plant biosynthetic enzyme selected from the group consisting of aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine synthase and cystathionine β -lyase, operably linked to regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the biosynthetic enzyme in the transformed host cell; (c) optionally purifying the biosynthetic enzyme expressed by the transformed host cell; (d) treating the biosynthetic enzyme with a compound to be tested; and (e) comparing the activity of the biosynthetic enzyme that has been treated with a test compound to the activity of an untreated biosynthetic enzyme.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE LISTINGS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts the biosynthetic pathway for the aspartate family of amino acids. The following abbreviations are used: AK = aspartokinase; ASADH = aspartic semialdehyde dehydrogenase; DHDPS = dihydrodipicolinate synthase;

DHDPR = dihydrodipicolinate reductase; DAPEP = diaminopimelate epimerase;
DAPDC = diaminopimelate decarboxylase; HDH = homoserine dehydrogenase;
HK = homoserine kinase; TS = threonine synthase; TD = threonine deaminase;
CγS = cystathionine γ-synthase; CβL = cystathionine β-lyase; MS = methionine synthase;
5 CS = cysteine synthase; and SAMS = S-adenosylmethionine synthase.

Figures 2 through 6 show the amino acid sequence alignments between the known art
sequences for aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase,
homoserine kinase, cysteine γ synthase, and cystathione β-lyase with the sequences included
in this application. Alignments were performed using the Clustal alogarithm described in
10 Higgins and Sharp (1989) (CABIOS 5:151-153). Amino acids conserved among all
sequences are indicated by an asterisk (*) above the alignment. Dashes are used by the
program to maximize the alignment. A description of Figures 2 through 6 follows:

Figure 2 shows a comparison of the aspartic semialdehyde dehydrogenase amino acid
sequences from corn contig assembled from clones p0003.cgpha22r: fis, cpe1c.pk009.b24,
15 p0016.ctscp83r, and p0075.cslab16r (SEQ ID NO:43), rice clone rlr48.pk0003.d12 (SEQ ID
NO:2), the contig of 5' RACE PCR and rice clone rlr48.pk0003.d12 (SEQ ID NO:45),
soybean clones sfl1.pk0122.f9 (SEQ ID NO:6), ses9c.pk001.a15: fis (SEQ ID NO:47), and
sfl1.pk0122.f9: fis (SEQ ID NO:49), wheat clones wr1.pk0004.c11 (SEQ ID NO:4) and
wdk1c.pk014.n5: fis (SEQ ID NO:51) with the *Legionella pneumophila* (NCBI General
20 Identifier No. 2645882; SEQ ID NO:7) and the *Aquifex aeolicus* sequences (NCBI General
Identifier No. 6225258; SEQ ID NO:52). Figure 2A: positions 1 through 120; Figure 2B:
positions 121 through 240; Figure 2C: positions 241 through 360; Figure 2D: positions 361
through 392.

Figure 3 shows a comparison of the diaminopimelate decarboxylase amino acid
sequences derived from corn clones cen3n.pk0067.a3 (SEQ ID NO:9) and cr1n.pk0103.d8
25 (SEQ ID NO:11), rice clone rl0n.pk0013.b9 (SEQ ID NO:13), soybean clones sr1.pk0132.c1
(SEQ ID NO:15), sdp3c.pk001.o15 (SEQ ID NO:19) and sdp3c.pk001.o15: fis (SEQ ID
NO:54), wheat clones wlk1.pk0012.c2 (SEQ ID NO:17) and wlk1.pk0012.c2: fis (SEQ ID
NO:56) with the *Pseudomonas aeruginosa* (NCBI General Identifier No. 118304; SEQ ID
30 NO:20) and *Arabidopsis thaliana* sequences (NCBI General Identifier No. 9279586; SEQ ID
NO:57). Figure 3A: positions 1 through 120; Figure 3B: positions 121 through 240;
Figure 3C: positions 241 through 360; Figure 3D: positions 361 through 480; Figure 3E:
positions 481 through 535.

Figure 4 shows a comparison of the homoserine kinase amino acid sequences derived
35 from corn clone cr1n.pk0009.g4 (SEQ ID NO:22), rice clones rca1c.pk005.k3 (SEQ ID
NO:24) and rca1c.pk005.k3: fis (SEQ ID NO:59), soybean clone ses8w.pk0020.b5 (SEQ ID
NO:26), wheat clone wl1n.pk0065.f2 (SEQ ID NO:28) with the *Methanococcus jannaschii*

(NCBI General Identifier No. 1591748; SEQ ID NO:29) and the *Arabidopsis thaliana* sequences (NCBI General Identifier No. 4927412; SEQ ID NO:60). Figure 4A: positions 1 through 180; Figure 4B: positions 181 through 360; Figure 4C: positions 361 through 396.

Figure 5 shows a comparison of the cysteine γ synthase amino acid sequences derived from the corn contig assembled from clones cco1n.pk083.j4, chp2.pk0016.b1, cpd1c.pk004.b20, cr1n.pk0083.c5, csi1.pk0003.g6, and p0126.cnlcb49r (SEQ ID NO:62), rice clone rls6.pk0068.b7: fis (SEQ ID NO:64), soybean clone se3.05h06 (SEQ ID NO:31) with the *Citrullus lanatus* sequence (NCBI General Identifier No. 540497; SEQ ID NO:32), the *Spinacia oleracea* sequence (NCBI General Identifier No. 540497; SEQ ID NO:65), and the *Solanum tuberosum* sequence (NCBI General Identifier No. 11131628; SEQ ID NO:66). Figure 5A: positions 1 through 180; Figure 5B: positions 181 through 360; Figure 5C: positions 361 through 424.

Figure 6 shows a comparison of the amino acid sequences of the cystathionine β -lyase derived from corn clone cen1.pk0061.d4 (SEQ ID NO:34), corn contig assembled from clones p0005.cbmei71r, p0014.ctuui39r, p0109.cdadg47r, and p0125.czaay16r (SEQ ID NO:68), rice clone rlr12.pk0026.g1 (SEQ ID NO:36), the contig of 5' PCR and rice clone rlr12.pk0026.g1: fis (SEQ ID NO:70), soybean clone sfl1.pk0012.c4 (SEQ ID NO:38), and wheat clones wr1.pk0091.g6 (SEQ ID NO:40) and wr1.pk0091.g6: fis (SEQ ID NO:72) with the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 1708993; SEQ ID NO:41). Figure 6A: positions 1 through 120; Figure 6B: positions 121 through 240; Figure 6C: positions 241 through 360; Figure 6D: positions 361 through 483.

Table 1 lists the polypeptides that are described herein, the designation of the cDNA clones that comprise the nucleic acid fragments encoding polypeptides representing all or a substantial portion of these polypeptides, and the corresponding identifier (SEQ ID NO:) as used in the attached Sequence Listing. The sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

TABLE 1
Plant Biosynthetic Enzymes

Polypeptide	Clone	SEQ ID NO:	
		(Nucleotide)	(Amino Acid)
rice ASADH	rlr48.pk0003.d12	1	2
wheat ASADH	wr1.pk0004.c11	3	4
soybean ASADH	sfl1.pk0122.f9	5	6
<i>L. pneumophila</i> ASADH	NCBI GI 2645882		7
corn DAPEP	cen3n.pk0067.a3	8	9
corn DAPEP	cr1n.pk0103.d8	10	11

rice DAPEP	rl0n.pk0013.b9	12	13
soybean DAPEP	sr1.pk0132.c1	14	15
wheat DAPEP	wlk1.pk0012.c2	16	17
soybean DAPEP	sdp3c.pk001.o15	18	19
<i>P. aeruginosa</i> DAPEP	NCBI GI 118304		20
corn HK	cr1n.pk0009.g4	21	22
rice HK	rca1c.pk005.k3	23	24
soybean HK	ses8w.pk0020.b5	25	26
wheat HK	wl1n.pk0065.f2	27	28
<i>M. jannaschii</i> HK	NCBI GI 1591748		29
soybean CγS	se3.05h06	30	31
<i>C. lanatus</i> CγS	NCBI GI 540497		32
corn CβL	cen1.pk0061.d4	33	34
rice CβL	rlr12.pk0026.g1	35	36
soybean CβL	sfl1.pk0012.c4	37	38
wheat CβL	wr1.pk0091.g6	39	40
<i>A. thaliana</i> CβL	NCBI GI 1708993		41
corn ASADH	Contig of:	42	43
	p0003.cgpha22r:fis		
	cpe1c.pk009.b24		
	p0016.ctscp83r		
	p0075.cslab16r		
rice ASADH	5' RACE PCR+	44	45
	rlr48.pk0003.d12		
soybean ASADH	ses9c.pk001.a15:fis	46	47
soybean ASADH	sfl1.pk0122.f9:fis	48	49
wheat ASADH	wdk1c.pk014.n5:fis	50	51
<i>A. aeolicus</i> ASADH	NCBI GI 6225258		52
soybean DAPEP	sdp3c.pk001.o15:fis	53	54
wheat DAPEP	wlk1.pk0012.c2:fis	55	56
<i>A. thaliana</i> DAPEP	NCBI GI 9279586		57
rice HK	rca1c.pk005.k3:fis	58	59
<i>A. thaliana</i> HK	NCBI GI 4927412		60
corn CγS	Contig of:	61	62
	cco1n.pk083.j4		
	chp2.pk0016.b1		
	cpd1c.pk004.b20		
	cr1n.pk0083.c5		
	csi1.pk0003.g6		
	p0126.cnlcb49r		
rice CγS	rls6.pk0068.b7:fis	63	64
<i>S. oleracea</i> CγS	NCBI GI 416869		65

<i>S. tuberosum</i> CγS	NCBI GI 11131628		66
corn CβL	Contig of: p0005.cbmei71r p0014.ctuui39r p0109.cdadg47r p0125.czaay16r	67	68
rice CβL	5'RACE PCR + rlr12.pk0026.g1:fis	69	70
wheat CβL	wr1.pk0091.g6:fis	71	72

The nucleotide and amino acid sequences shown in SEQ ID NOs:1 through 41 are found, with the same SEQ ID NO, in U.S. Application No. 09/424,976. All or a portion of some of the sequences in the present application are found in the provisional applications for which the present application claims priority to. Table 1A indicates the SEQ ID NO: in the present application and the corresponding SEQ ID NO: in the previously-filed provisional application.

TABLE 1A
Sequence Priority

Application No. 09/424,976	Provisional Application No. 60/049406	Provisional Application No. 60/065385
SEQ ID NO:1	SEQ ID NO:1	
SEQ ID NO:2	SEQ ID NO:2	
SEQ ID NO:3	SEQ ID NO:3*	
SEQ ID NO:4	SEQ ID NO:4*	
SEQ ID NO:8	SEQ ID NO:7	SEQ ID NO:8
SEQ ID NO:9	SEQ ID NO:8	SEQ ID NO:9
SEQ ID NO:12	SEQ ID NO:9	
SEQ ID NO:13	SEQ ID NO:10	
SEQ ID NO:14	SEQ ID NO:11	SEQ ID NO:5
SEQ ID NO:15	SEQ ID NO:12	SEQ ID NO:6
SEQ ID NO:21	SEQ ID NO:13	SEQ ID NO:10*
SEQ ID NO:22	SEQ ID NO:14	SEQ ID NOs:11* and 14*
SEQ ID NO:23	SEQ ID NO:17*	SEQ ID NO:15
SEQ ID NO:24	SEQ ID NO:18*	SEQ ID NO:16
SEQ ID NO:25	SEQ ID NO:15	SEQ ID NO:13
SEQ ID NO:26	SEQ ID NO:16	SEQ ID NO:14
SEQ ID NO:30	SEQ ID NO:19	SEQ ID NO:17
SEQ ID NO:31	SEQ ID NO:20	SEQ ID NO:18
SEQ ID NO:33*	SEQ ID NO:21	SEQ ID NO:19
SEQ ID NO:34	SEQ ID NO:22	SEQ ID NO:20
SEQ ID NO:37	SEQ ID NO:23	SEQ ID NO:21*
SEQ ID NO:38	SEQ ID NO:24	SEQ ID NO:22*

* Indicates that only a portion of the sequence was in the application.

5 The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Res.* 13:3021-3030 (1985) and in the *Biochemical J.* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

10 In the context of this disclosure, a number of terms shall be utilized. The terms "polynucleotide," "polynucleotide sequence," "nucleic acid sequence," and "nucleic acid fragment"/"isolated nucleic acid fragment" are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of RNA or DNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof. An isolated polynucleotide of the present invention may include at least 15 30 contiguous nucleotides, preferably at least 40 contiguous nucleotides, most preferably at least 60 contiguous nucleotides derived from SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, SEQ ID NOs:8, 10, 12, 14, 16, 18, 53 and 55, SEQ ID NOs:21, 23, 25, 27, and 58, SEQ ID NOs:30, 61, and 63, and SEQ ID NOs:33, 35, 37, 39, 67, 69, and 71, or the complement of 20 such sequences.

25 The term "isolated" polynucleotide refers to a polynucleotide that is substantially free from other nucleic acid sequences with which it is normally associated such as other chromosomal and extrachromosomal DNA and RNA. Isolated polynucleotides may be purified from a host cell in which they naturally occur. Conventional nucleic acid purification methods known to skilled artisans may be used to obtain isolated polynucleotides. The term also embraces recombinant polynucleotides and chemically synthesized polynucleotides.

30 The term "recombinant" means, for example, that a nucleic acid sequence is made by an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis or by the manipulation of isolated nucleic acids by genetic engineering techniques.

35 As used herein, "contig" refers to a nucleotide sequence that is assembled from two or more constituent nucleotide sequences that share common or overlapping regions of sequence homology. For example, the nucleotide sequences of two or more nucleic acid fragments can be compared and aligned in order to identify common or overlapping sequences. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences (and thus their corresponding nucleic acid fragments) can be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the polypeptide encoded by the nucleotide sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by gene silencing through for example antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate gene silencing or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary nucleotide or amino acid sequences and includes functional equivalents thereof. The terms "substantially similar" and "corresponding substantially" are used interchangeably herein.

Substantially similar nucleic acid fragments may be selected by screening nucleic acid fragments representing subfragments or modifications of the nucleic acid fragments of the instant invention, wherein one or more nucleotides are substituted, deleted and/or inserted, for their ability to affect the level of the polypeptide encoded by the unmodified nucleic acid fragment in a plant or plant cell. For example, a substantially similar nucleic acid fragment representing at least 30 contiguous nucleotides derived from the instant nucleic acid fragment can be constructed and introduced into a plant or plant cell. The level of the polypeptide encoded by the unmodified nucleic acid fragment present in a plant or plant cell exposed to the substantially similar nucleic fragment can then be compared to the level of the polypeptide in a plant or plant cell that is not exposed to the substantially similar nucleic acid fragment.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by using nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a nucleic acid fragment which result in the production of a chemically equivalent amino acid at a given site, but do not affect the functional properties of the encoded polypeptide, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which

result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Consequently, an isolated polynucleotide comprising a nucleotide sequence of at least 30 (preferably at least 40, most preferably at least 60) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, SEQ ID NOs:8, 10, 12, 14, 16, 18, 53 and 55, SEQ ID NOs:21, 23, 25, 27, and 58, SEQ ID NOs:30, 61, and 63, and SEQ ID NOs:33, 35, 37, 39, 67, 69, and 71 and the complement of such nucleotide sequences may be used in methods of selecting an isolated polynucleotide that affects the expression of an aspartic-semialdehyde dehydrogenase, a diaminopimelate decarboxylase, a homoserine kinase, a cysteine γ synthase, or a cystathionine β -lyase polypeptide in a host cell. A method of selecting an isolated polynucleotide that affects the level of expression of a polypeptide in a host cell may comprise the steps of: constructing an isolated polynucleotide of the present invention or an isolated chimeric gene of the present invention; introducing the isolated polynucleotide or the isolated chimeric gene into a host cell; measuring the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide; and comparing the level of a polypeptide or enzyme activity in the host cell containing the isolated polynucleotide with the level of a polypeptide or enzyme activity in a host cell that does not contain the isolated polynucleotide.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize. Estimates of such homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 min, then repeated with 2X SSC, 0.5% SDS at 45°C for 30 min, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 min. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the

amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Suitable nucleic acid fragments (isolated polynucleotides of the present invention) encode polypeptides that are at least about 70% identical, preferably at least about 80% identical to the amino acid sequences reported herein. Preferred nucleic acid fragments encode amino acid sequences that are about 85% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are at least about 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are at least about 95% identical to the amino acid sequences reported herein. Suitable nucleic acid fragments not only have the above identities but typically encode a polypeptide having at least 50 amino acids, preferably at least 100 amino acids, more preferably at least 150 amino acids, still more preferably at least 200 amino acids, and most preferably at least 250 amino acids. Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises an amino acid or a nucleotide sequence that is sufficient to afford putative identification of the protein or gene that the amino acid or nucleotide sequence comprises. Amino acid and nucleotide sequences can be evaluated either manually, by one skilled in the art, or by using computer-based sequence comparison and identification tools that employ algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more contiguous nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene-specific oligonucleotide probes comprising 30 or more contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12 or more nucleotides may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises a nucleotide sequence that will afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches amino acid and nucleotide sequences encoding polypeptides that

comprise one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without affecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment comprising a nucleotide sequence that encodes all or a substantial portion of the amino acid sequences set forth herein. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a nucleic acid fragment for improved expression in a host cell, it is desirable to design the nucleic acid fragment such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic nucleic acid fragments" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form larger nucleic acid fragments which may then be enzymatically assembled to construct the entire desired nucleic acid fragment. "Chemically synthesized", as related to a nucleic acid fragment, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of nucleic acid fragments may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the nucleic acid fragments can be tailored for optimal gene expression based on optimization of the nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

"Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign-gene" refers to a gene not normally found in the host organism, but

that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a nucleotide sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a nucleotide sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a nucleotide sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or may be composed of different elements derived from different promoters found in nature, or may even comprise synthetic nucleotide segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a nucleic acid fragment to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, nucleic acid fragments of different lengths may have identical promoter activity.

“Translation leader sequence” refers to a nucleotide sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner and Foster (1995) *Mol. Biotechnol.* 3:225-236).

“3' non-coding sequences” refer to nucleotide sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid

tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into polypeptides by the cell. “cDNA” refers to DNA that is complementary to and derived from an mRNA template. The cDNA can be single-stranded or converted to double stranded form using, for example, the Klenow fragment of DNA polymerase I. “Sense-RNA” refers to an RNA transcript that includes the mRNA and so can be translated into a polypeptide by the cell. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (see U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific nucleotide sequence, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term “operably linked” refers to the association of two or more nucleic acid fragments on a single polynucleotide so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term “expression”, as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. “Antisense inhibition” refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. “Co-suppression” refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

A “protein” or “polypeptide” is a chain of amino acids arranged in a specific order determined by the coding sequence in a polynucleotide encoding the polypeptide. Each protein or polypeptide has a unique function.

“Altered levels” or “altered expression” refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

“Mature protein” or the term “mature” when used in describing a protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. “Precursor protein” or the term “precursor” when used in describing a protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A “chloroplast transit peptide” is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. “Chloroplast transit sequence” refers to a nucleotide sequence that encodes a chloroplast transit peptide. A “signal peptide” is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

“Transformation” refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or “gene gun” transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference). Thus, isolated polynucleotides of the present invention can be incorporated into recombinant constructs, typically DNA constructs, capable of introduction into and replication in a host cell. Such a construct can be a vector that includes a replication system and sequences that are capable of transcription and translation of a polypeptide-encoding sequence in a given host cell. A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described in, e.g., Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985, supp. 1987; Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, 1989; and Flevin et al., *Plant Molecular Biology Manual*, Kluwer Academic Publishers, 1990. Typically, plant expression vectors include, for example, one or more cloned plant genes under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable

marker. Such plant expression vectors also can contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook et al. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

"PCR" or "polymerase chain reaction" is well known by those skilled in the art as a technique used for the amplification of specific DNA segments (U.S. Patent Nos. 4,683,195 and 4,800,159).

The present invention concerns isolated polynucleotides comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51; (b) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:9, 11, 13, 15, 17, 19, 54 and 56; (c) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 80% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:22, 24, 26, 28, and 59; (d) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 95% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:31, 62, and 64; and (e) a nucleotide sequence encoding a polypeptide of at least 60 amino acids having at least 85% identity based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72. It is preferred that the identity be at least 85%, it is preferable if the identity is at least 90%, it is more preferred that the identity be at least 95%. This invention also relates to the isolated complement of such polynucleotides, wherein the complement and the polynucleotide consist of the same number of nucleotides, and the nucleotide sequences of the complement and the polynucleotide have 100% complementarity.

Preferably, the isolated polynucleotide of the claimed invention comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, 8, 10, 12, 14, 16, 18, 53, 55, 21, 23, 25, 27, 58, 30, 61, 63, 33, 35, 37, 39, 67, 69, and 71.

Nucleic acid fragments encoding at least a portion of several plant amino acid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other aspartic semialdehyde dehydrogenases, diaminopimelate decarboxylases, homoserine kinases, cysteine γ synthases or cystathionine β -lyases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, an entire sequence can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl.*

Acad. Sci. USA 86:5673-5677; Loh et al. (1989) *Science* 243:217-220). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman and Martin (1989) *Techniques* 1:165). Consequently, a polynucleotide comprising a nucleotide sequence of at least 30 (preferably at least 40, most preferably at least 60) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, 8, 10, 12, 14, 16, 18, 59, 61, 21, 23, 25, 27, 64, 30, 33, 35, 37, 39, 53, 55, and 57 and the complement of such nucleotide sequences may be used in such methods to obtain a nucleic acid fragment encoding a substantial portion of an amino acid sequence of a polypeptide.

The present invention relates to a method of obtaining a nucleic acid fragment encoding a substantial portion of an aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine synthase, or cystathionine β -lyase polypeptide, preferably a substantial portion of a plant aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine synthase, or cystathionine β -lyase polypeptide, comprising the steps of: synthesizing an oligonucleotide primer comprising a nucleotide sequence of at least 30 (preferably at least 40, most preferably at least 60) contiguous nucleotides derived from a nucleotide sequence selected from the group consisting of SEQ ID NOs:1, 3, 5, 42, 44, 46, 48, 50, 8, 10, 12, 14, 16, 18, 53, 55, 21, 23, 25, 27, 58, 30, 61, 63, 33, 35, 37, 39, 67, 69, and 71, and the complement of such nucleotide sequences; and amplifying a nucleic acid fragment (preferably a cDNA inserted in a cloning vector) using the oligonucleotide primer. The amplified nucleic acid fragment preferably will encode a portion of an aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine synthase, or cystathionine β -lyase polypeptide.

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner (1984) *Adv. Immunol.* 36:1-34; Maniatis).

In another embodiment, this invention concerns viruses and host cells comprising either the chimeric genes of the invention as described herein or an isolated polynucleotide of the invention as described herein. Examples of host cells which can be used to practice the invention include, but are not limited to, yeast, bacteria, and plants.

As was noted above, the nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed polypeptides are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally

found. This would have the effect of altering the level of free amino acids in those cells. Specifically, the enzymes of the present invention form part of the pathway towards the biosynthesis of lysine, threonine, methionine, cysteine and isoleucine. In particular, altering the level and/or function of cystathionine beta-lyase will result in changes in the rate of methionine biosynthesis. Altering the level and/or function of diaminopimelate decarboxylase will result in changes in the rate of lysine biosynthesis. Altering the level and/or function of aspartate-semialdehyde dehydrogenase will result in changes in the lysine, methionine, or threonine content, especially in wheat. Altering the level of cysteine γ synthase will result in changes in the rate of cysteine and/or methionine biosynthesis; using this gene it will also be possible to control sulfur metabolism. Altering the level of homoserine kinase may be used to regulate threonine and methionine levels. Polypeptides encoding at least a portion of aspartic semialdehyde dehydrogenase, diaminopimelate decarboxylase, homoserine kinase, cysteine synthase, or cystathionine β -lyase may also be used in herbicide identification and design.

Overexpression of the proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. The chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant isolated polynucleotide (or chimeric gene) may be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al. (1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant polypeptides to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by directing the coding sequence to encode the instant polypeptides with appropriate intracellular targeting sequences such as transit sequences (Keegstra (1989) *Cell* 56:247-253), signal sequences or

sequences encoding endoplasmic reticulum localization (Chrispeels (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel (1992) *Plant Phys.* 100:1627-1632) with or without removing targeting sequences that are already present. While the references cited give examples of each of these, the list is not exhaustive and more
5 targeting signals of use may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding the instant polypeptides in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant polypeptide can be constructed by linking a gene or gene fragment encoding that polypeptide to plant promoter sequences.

10 Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

15 Molecular genetic solutions to the generation of plants with altered gene expression have a decided advantage over more traditional plant breeding approaches. Changes in plant phenotypes can be produced by specifically inhibiting expression of one or more genes by antisense inhibition or cosuppression (U.S. Patent Nos. 5,190,931, 5,107,065 and 5,283,323). An antisense or cosuppression construct would act as a dominant negative
20 regulator of gene activity. While conventional mutations can yield negative regulation of gene activity these effects are most likely recessive. The dominant negative regulation available with a transgenic approach may be advantageous from a breeding perspective. In addition, the ability to restrict the expression of a specific phenotype to the reproductive tissues of the plant by the use of tissue specific promoters may confer agronomic
25 advantages relative to conventional mutations which may have an effect in all tissues in which a mutant gene is ordinarily expressed.

The person skilled in the art will know that special considerations are associated with the use of antisense or cosuppression technologies in order to reduce expression of particular genes. For example, the proper level of expression of sense or antisense genes may require
30 the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan. Once transgenic plants are obtained by one of the methods described above, it will be necessary to screen individual transgenics for those that most effectively display the desired phenotype. Accordingly, the skilled artisan will develop methods for screening large numbers of transformants. The nature of these screens will generally be chosen on
35 practical grounds. For example, one can screen by looking for changes in gene expression by using antibodies specific for the protein encoded by the gene being suppressed, or one could establish assays that specifically measure enzyme activity. A preferred method will be

one which allows large numbers of samples to be processed rapidly, since it will be expected that a large number of transformants will be negative for the desired phenotype.

In another embodiment, the present invention concerns an aspartic-semialdehyde dehydrogenase polypeptide of at least 50 amino acids comprising at least 70% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51, a diaminopimelate decarboxylase polypeptide of at least 60 amino acids comprising at least 95% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:9, 11, 13, 15, 17, 19, 60, and 62, a homoserine kinase polypeptide of at least 60 amino acids comprising at least 70% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:22, 24, 26, 28, and 65, a cysteine synthase polypeptide of at least 60 amino acids comprising at least 90% identity based on the Clustal method of alignment compared to a polypeptide of SEQ ID NO:31, or a cystathionine β -lyase polypeptide of at least 60 amino acids comprising at least 85% identity based on the Clustal method of alignment compared to a polypeptide selected from the group consisting of SEQ ID NOs:34, 36, 38, 40, 54, 56, and 58.

The instant polypeptides (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting the polypeptides of the instant invention *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant polypeptides are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant polypeptides. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded plant biosynthetic enzymes. An example of a vector for high level expression of the instant polypeptides in a bacterial host is provided (Example 10).

Additionally, the instant polypeptides can be used as a target to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the polypeptides described herein catalyze various steps in a pathway leading to production of several essential amino acids. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition of plant growth. Thus, the instant polypeptides could be appropriate for new herbicide discovery and design.

All or a substantial portion of the polynucleotides of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of,

and used as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein et al. (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in Bernatzky and Tanksley (1986) *Plant Mol. Biol. Reporter* 4:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan et al. (1995) *Genome Res.* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian (1989) *J. Lab. Clin. Med.* 11:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter et al. (1997) *Nat. Genet.* 7:22-28) and Happy Mapping (Dear and Cook (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension

5 reactions. The design of such primers is well known to those skilled in the art. In methods
employing PCR-based genetic mapping, it may be necessary to identify DNA sequence
differences between the parents of the mapping cross in the region corresponding to the
instant nucleic acid sequence. This, however, is generally not necessary for mapping
methods.

10 Loss of function mutant phenotypes may be identified for the instant cDNA clones
either by targeted gene disruption protocols or by identifying specific mutants for these
genes contained in a maize population carrying mutations in all possible genes (Ballinger
and Benzer (1989) *Proc. Natl. Acad. Sci USA* 86:9402-9406; Koes et al. (1995) *Proc. Natl.*
15 *Acad. Sci USA* 92:8149-8153; Bensen et al. (1995) *Plant Cell* 7:75-84). The latter approach
may be accomplished in two ways. First, short segments of the instant nucleic acid
fragments may be used in polymerase chain reaction protocols in conjunction with a
mutation tag sequence primer on DNAs prepared from a population of plants in which
Mutator transposons or some other mutation-causing DNA element has been introduced (see
20 Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates
the insertion of the mutation tag element in or near the plant gene encoding the instant
polypeptides. Alternatively, the instant nucleic acid fragment may be used as a
hybridization probe against PCR amplification products generated from the mutation
population using the mutation tag sequence primer in conjunction with an arbitrary genomic
25 site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With
either method, a plant containing a mutation in the endogenous gene encoding the instant
polypeptides can be identified and obtained. This mutant plant can then be used to
determine or confirm the natural function of the instant polypeptides disclosed herein.

EXAMPLES

25 The present invention is further defined in the following Examples, in which parts
and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be
understood that these Examples, while indicating preferred embodiments of the invention,
are given by way of illustration only. From the above discussion and these Examples, one
skilled in the art can ascertain the essential characteristics of this invention, and without
30 departing from the spirit and scope thereof, can make various changes and modifications of
the invention to adapt it to various usages and conditions. Thus, various modifications of the
invention in addition to those shown and described herein will be apparent to those skilled in
the art from the foregoing description. Such modifications are also intended to fall within
the scope of the appended claims.

35 The disclosure of each reference set forth herein is incorporated herein by reference in
its entirety.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean, and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean, and Wheat

Library	Tissue	Clone
cen1	Corn Endosperm 12 Days After Pollination	cen1.pk0061.d4
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0067.a3
cpe1c	Corn pooled BMS treated with chemicals related to phosphatase**	cpe1c.pk009.b24
cr1n	Corn Root From 7 Day Seedlings*	cr1n.pk0009.g4
cr1n	Corn Root From 7 Day Seedlings*	cr1n.pk0103.d8
p0003	Corn Premeiotic Ear Shoot , 0.2-4 cm	p0003.cgpha22r:fis
p0005	Corn Immature Ear	p0005.cbmei71r
p0014	Corn Leaves 7 and 8 from Plant Transformed with G-protein Gene, <i>C. heterostrophus</i> Resistant	p0014.ctuui39r
p0016	Corn Tassel Shoots (0.1-1.4 cm), Pooled	p0016.ctscp83r
p0075	Corn Shoot And Leaf Material From Dark-Grown 7 Day-Old Seedlings	p0075.cslab16r
p0109	Corn Leaves From Les9 Transition Zone and Les9 Mature Lesions, Pooled***	p0109.cdadg47r
p0125	Corn Anther Prophase I*	p0125.czaay16r
rca1c	Rice Nipponbare Callus	rca1c.pk005.k3
rl0n	Rice Leaf 15 Days After Germination*	rl0n.pk0013.b9
rlr12	Rice Leaf 15 Days After Germination, 12 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO)	rlr12.pk0026.g1
rlr48	Rice Leaf 15 Days After Germination 48 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO)	rlr48.pk0003.d12
se3	Soybean Embryo 13 Days After Flowering	sdp3c.pk001.o15
sdp3c	Soybean Developing Pods 8-9 mm	se3.05h06
ses8w	Mature Soybean Embryo 8 Weeks After Subculture	ses8w.pk0020.b5
ses9c	Soybean Embryogenic Suspension	ses9c.pk001.a15:fis
sfl1	Soybean Immature Flower	sfl1.pk0012.c4
sfl1	Soybean Immature Flower	sfl1.pk0122.f9
sr1	Soybean Root From 10 Day Old Seedlings	sr1.pk0132.c1
wdk1c	Wheat Developing Kernel, 3 Days After Anthesis	wdk1c.pk014.n5:fis

wl1n	Wheat Leaf from 7 Day Old Seedling*	wl1n.pk0065.f2
wlk1	Wheat Seedlings 1 hour After Fungicide Treatment****	wlk1.pk0012.c2
wr1	Wheat Root From 7 Day Old Seedlings	wr1.pk0004.c11
wr1	Wheat Root From 7 Day Old Seedlings	wr1.pk0091.g6

* These libraries were normalized essentially as described in U.S. Patent No. 5,482,845.

** Chemicals used included okadaic acid, cyclosporin A, calyculin A, and cypermethrin, all of which are commercially available from Molecular Biology supply sources including Calbiochem-Novabiochem Corp.

5 *** Les9 mutants reviewed in "An update on lesion mutants" Hoisington (1986) Maize Genetic Coop. News Lett. 60: 50-51.

**** Application of 6-iodo-2-propoxy-3-propyl-4(3*H*)-quinazolinone; synthesis and methods of using this compound are described in USSN 08/545,827, incorporated herein by reference.

10

036154560

25

30

35

cDNA libraries may be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams et al., (1991) *Science* 252:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

Full-insert sequence (FIS) data is generated utilizing a modified transposition protocol. Clones identified for FIS are recovered from archived glycerol stocks as single colonies, and plasmid DNAs are isolated via alkaline lysis. Isolated DNA templates are reacted with vector primed M13 forward and reverse oligonucleotides in a PCR-based sequencing reaction and loaded onto automated sequencers. Confirmation of clone identification is performed by sequence alignment to the original EST sequence from which the FIS request is made.

Confirmed templates are transposed via the Primer Island transposition kit (PE Applied Biosystems, Foster City, CA) which is based upon the *Saccharomyces cerevisiae*

Ty1 transposable element (Devine and Boeke (1994) *Nucleic Acids Res.* 22:3765-3772). The *in vitro* transposition system places unique binding sites randomly throughout a population of large DNA molecules. The transposed DNA is then used to transform DH10B electro-competent cells (Gibco BRL/Life Technologies, Rockville, MD) via electroporation. The transposable element contains an additional selectable marker (named DHFR; Fling and Richards (1983) *Nucleic Acids Res.* 11:5147-5158), allowing for dual selection on agar plates of only those subclones containing the integrated transposon. Multiple subclones are randomly selected from each transposition reaction, plasmid DNAs are prepared via alkaline lysis, and templates are sequenced (ABI Prism dye-terminator ReadyReaction mix) outward from the transposition event site, utilizing unique primers specific to the binding sites within the transposon.

Sequence data is collected (ABI Prism Collections) and assembled using Phred/Phrap (P. Green, University of Washington, Seattle). Phrep/Phrap is a public domain software program which re-reads the ABI sequence data, re-calls the bases, assigns quality values, and writes the base calls and quality values into editable output files. The Phrap sequence assembly program uses these quality values to increase the accuracy of the assembled sequence contigs. Assemblies are viewed by the Consed sequence editor (D. Gordon, University of Washington, Seattle).

EXAMPLE 2

Identification of cDNA Clones

cDNA clones encoding plant amino acid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al. (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish and States (1993) *Nat. Genet.* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

ESTs submitted for analysis are compared to the genbank database as described above. ESTs that contain sequences more 5- or 3-prime can be found by using the BLASTn algorithm (Altschul et al (1997) *Nucleic Acids Res.* 25:3389-3402.) against the DuPont proprietary database comparing nucleotide sequences that share common or overlapping regions of sequence homology. Where common or overlapping sequences exist between two or more nucleic acid fragments, the sequences can be assembled into a single contiguous nucleotide sequence, thus extending the original fragment in either the 5 or 3 prime direction. Once the most 5-prime EST is identified, its complete sequence can be determined by Full Insert Sequencing as described in Example 1. Homologous genes belonging to different species can be found by comparing the amino acid sequence of a known gene (from either a proprietary source or a public database) against an EST database using the tBLASTn algorithm. The tBLASTn algorithm searches an amino acid query against a nucleotide database that is translated in all 6 reading frames. This search allows for differences in nucleotide codon usage between different species, and for codon degeneracy.

EXAMPLE 3

Characterization of cDNA Clones Encoding Aspartate Semialdehyde Dehydrogenase

The BLASTX search using the EST sequences from clones listed in Table 3 revealed similarity of the polypeptides encoded by the cDNAs to aspartate semialdehyde dehydrogenase from *Synechocystis sp.* (DDJB Accession No. D64006; NCBI General Identifier No. 1001379) or *Legionella pneumophila* (GenBank Accession No. AF034213; NCBI General Identifier No. 2645882). Shown in Table 3 are the BLAST results for individual ESTs ("EST"), or for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 3
BLAST Results for Sequences Encoding Polypeptides Homologous
to Aspartate Semialdehyde Dehydrogenase

Clone	Status	BLAST pLog Score	
		<i>Synechocystis sp.</i> GI 1001379	<i>Legionella pneumophila</i> GI 2645882
rlr48.pk0003.d12	FIS	51.00	36.00
wr1.pk0004.c11	EST	67.96	44.74
sfl1.pk0122.f9	EST		6.60

The sequence of the entire cDNA insert in clone sfl1.pk0122.f9 was determined, RACE PCR was used to obtain the 5' portion of the rice aspartate semialdehyde dehydrogenase, and further sequencing and searching of the DuPont proprietary database allowed the identification of a corn and other a soybean, and wheat clones encoding

aspartate semialdehyde dehydrogenase. The BLASTX search using the EST sequences from clones listed in Table 4 revealed similarity of the polypeptides encoded by the cDNAs to aspartate semialdehyde dehydrogenase from *Aquifex aeolicus* (NCBI General Identifier No. 6225258). Shown in Table 4 are the BLAST results for the sequences of contigs assembled from two or more ESTs ("Contig"), or the sequences encoding the entire protein derived from either the entire cDNA inserts comprising the indicated cDNA clones or contigs assembled from 5' RACE PCR and the sequence of the entire cDNA insert in the indicated cDNA clone ("CGS"):

TABLE 4

BLAST Results for Sequences Encoding Polypeptides Homologous to Aspartate Semialdehyde Dehydrogenase

Clone	Status	BLAST pLog Score <i>Aquifex aeolicus</i> GI 6225258
Contig of: cpe1c.pk009.b24 p0003.cgpha22r:fis p0016.ctscp83r p0075.cslab16r	Contig	78.70
5' RACE PCR + rlr48.pk0003.d12:fis	CGS	89.20
ses9c.pk001.a15:fis	CGS	87.40
sfl1.pk0122.f9:fis	CGS	88.10
wdk1c.pk014.n5:fis	CGS	91.50

Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51 with the *Legionella pneumophila* sequence (NCBI General Identifier No. 2645882; SEQ ID NO:7) and the *Aquifex aeolicus* sequence (NCBI General Identifier No. 6225258; SEQ ID NO:52). The data in Table 5 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 43, 45, 47, 49, and 51 with the *Legionella pneumophila* sequence (NCBI General Identifier No. 2645882; SEQ ID NO:7) and the *Aquifex aeolicus* sequence (NCBI General Identifier No. 6225258; SEQ ID NO:52).

TABLE 5

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Aspartate Semialdehyde Dehydrogenase

Clone	amino acid SEQ ID NO.	Percent Identity to	
		2645882	6225258
rlr48.pk0003.d12	2	42.1	45.6
wr1.pk0004.c11	4	42.3	44.8
sfl1.pk0122.f9	6	29.1	25.6
Contig of:	43	41.2	45.9
cpe1c.pk009.b24			
p0003.cgpha22r: fis			
p0016.ctscp83r			
p0075.cslab16r			
5' RACE PCR + rlr48.pk0003.d12: fis	45	43.2	47.0
ses9c.pk001.a15: fis	47	43.5	49.1
sfl1.pk0122.f9: fis	49	41.2	45.6
wdk1c.pk014.n5: fis	51	43.2	49.4

As seen in Figure 2, the amino acid sequence shown in SEQ ID NO:2 is identical to amino acids 181 through 375 of SEQ ID NO:45; the sequence shown in SEQ ID NO:4 is identical to amino acids 173 through 374 of the sequence shown in SEQ ID NO:51; the sequence shown in SEQ ID NO:6 is identical to amino acids 1 through 86 of the sequence shown in SEQ ID NO:49; there are 5 amino acid differences between the sequences shown in SEQ ID NO:47 and SEQ ID NO:49; there are 18 amino acid differences between amino acids 89 through 375 of the sequence shown in SEQ ID NO:43 and the sequence shown in SEQ ID NO:45; and there are 15 differences between the amino acid sequences shown in SEQ ID NO:45 and in SEQ ID NO:51.

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a corn aspartate semialdehyde dehydrogenase, a substantial portion and an entire rice aspartate semialdehyde dehydrogenase, a portion and an entire

wheat aspartate semialdehyde dehydrogenase, and a portion and an two entire soybean aspartate semialdehyde dehydrogenases.

EXAMPLE 4

Characterization of cDNA Clones Encoding Diaminopimelate Decarboxylase

The BLASTX search using the EST sequences from clones listed in Table 6 revealed similarity of the polypeptides encoded by the cDNAs to diaminopimelate decarboxylase from *Aquifex aeolicus* (GenBank Accession No. AE000728 and NCBI General Identifier No. 2983642) and *Pseudomonas aeruginosa* (GenBank Accession No. M23174 and NCBI General Identifier No. 118304). Shown in Table 6 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or the sequences of FISs encoding an entire protein ("CGS"):

TABLE 6

BLAST Results for Sequences Encoding Polypeptides Homologous to Diaminopimelate Decarboxylase

Clone	Status	BLAST pLog Score	
		GI 2983642 (<i>A. aeolicus</i>)	GI 118304 (<i>P. aeruginosa</i>)
cen3n.pk0067.a3	FIS	58.22	56.00
cr1n.pk0103.d8	CGS	75.25	79.12
rl0n.pk0013.b9	FIS	46.40	44.00
sr1.pk0132.c1	FIS	44.70	39.15
wlk1.pk0012.c2	EST	20.48	19.05

An additional soybean clone, sdp3c.pk001.o15, was identified as sharing homology with sr1.pk0132.c1. BLASTX search using the nucleotide sequences from clone sdp3c.pk001.o15 revealed similarity of the proteins encoded by the cDNA to diaminopimelate decarboxylase from *Pseudomonas fluorescens* (EMBO Accession No. Y12268; NCBI General Identifier No. 1929095). This EST yields a pLog value of 8.66 versus the *Pseudomonas fluorescens* sequence.

The sequence of the entire cDNA insert in clones sdp3c.pk001.o15 and wlk1.pk0012.c2 was determined. The BLASTX search using the EST sequences from clones listed in Table 7 revealed similarity of the polypeptides encoded by the cDNAs to diaminopimelate decarboxylase from *Aquifex aeolicus* (NCBI General Identifier No. 6225241) or by the *Arabidopsis thaliana* contig containing similarity with diaminopimelate decarboxylases (NCBI General Identifier No. 9279586). Shown in Table 7 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or the sequences of FISs encoding the entire protein ("CGS"):

TABLE 7

BLAST Results for Sequences Encoding Polypeptides Homologous
to Diaminopimelate Decarboxylase

Clone	Status	Homolog	BLAST pLog Score
sdp3c.pk001.o15:fis	CGS	GI 6225241 (<i>A. aeolicus</i>)	76.40
wlk1.pk0012.c2:fis	FIS	GI 9279586 (<i>A. thaliana</i>)	94.40

Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:9, 11, 13, 15, 17, 19, 54, and 56 with the *Pseudomonas aeruginosa* sequence (NCBI General Identifier No. 118304; SEQ ID NO:20) and the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 9279586, SEQ ID NO:57). The data in Table 8 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:9, 11, 13, 15, 17, 19, 54, and 56 with the *Pseudomonas aeruginosa* sequence (NCBI General Identifier No. 118304; SEQ ID NO:20) and the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 9279586; SEQ ID NO:57).

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide
Sequences of cDNA Clones Encoding Polypeptides Homologous
to Diaminopimelate Decarboxylase

Clone	Amino acid SEQ ID NO.	Percent Identity to	
		118304	9279586
cen3n.pk0067.a3	9	34.0	82.2
cr1n.pk0103.d8	11	35.9	70.6
rl0n.pk0013.b9	13	32.4	76.8
sr1.pk0132.c1	15	29.7	86.1
wlk1.pk0012.c2	17	42.5	93.2
sdp3c.pk001.o15	19	41.9	87.1
sdp3c.pk001.o15:fis	54	32.5	74.9
wlk1.pk0012.c2:fis	56	32.	84.9

The amino acid sequence set forth in SEQ ID NO:19 is identical to amino acids 112 through 173 of the amino acid sequence set forth in SEQ ID NO:54. The amino acid sequence set forth in SEQ IDNO:17 is identical to amino acids 24 through 96 of the amino acid sequence set forth in SEQ ID NO:56.

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default

parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of one corn, one rice, two soybean and one wheat diaminopimelate decarboxylases and entire corn and soybean diaminopimelate decarboxylases.

EXAMPLE 5

Characterization of cDNA Clones Encoding Homoserine Kinase

The BLASTX search using the EST sequences from clones listed in Table 9 revealed similarity of the polypeptides encoded by the cDNAs to homoserine kinase from *Methanococcus jannaschii* (GenBank Accession No. U67553 and NCBI General Identifier No. 1591748). Shown in Table 9 are the BLAST results for individual ESTs ("EST") or for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"):

TABLE 9

BLAST Results for Sequences Encoding Polypeptides Homologous to Homoserine Kinase

Clone	Status	BLAST pLog Score	
		GI 1591748 (<i>Methanococcus jannaschii</i>)	
cr1n.pk0009.g4	FIS	19.30	
rca1c.pk005.k3	EST	15.21	
ses8w.pk0020.b5	FIS	35.30	
wl1n.pk0065.f2	EST	5.68	

The sequence of the entire cDNA insert in clone rca1c.pk005.k3 was determined. The BLASTX search using the EST sequences from clones listed in Table 10 revealed similarity of the polypeptides encoded by the cDNAs to homoserine kinase from *Arabidopsis thaliana* (NCBI General Identifier No. 4927412). Shown in Table 10 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clone ("FIS"):

TABLE 10

BLAST Results for Sequences Encoding Polypeptides Homologous to Homoserine Kinase

Clone	Status	BLAST pLog Score	
		4927412 (<i>Arabidopsis thaliana</i>)	
rca1c.pk005.k3: fis	FIS	88.40	

Figure 4 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:22, 24, 26, 28, and 59 with the *Methanococcus jannaschii* sequence (NCBI General Identifier No. 1591748; SEQ ID NO:29) and the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 4927412; SEQ ID NO:60). The data in Table 11 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:22, 24, 26, 28, and 59 with the *Methanococcus jannaschii* sequence (NCBI General Identifier No. 1591748; SEQ ID NO:29) and the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 4927412; SEQ ID NO:60).

TABLE 11

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Homoserine Kinase

Clone	SEQ ID NO.	Percent Identity to	
		NCBI GI 1591748	NCBI GI 4927412
cr1n.pk0009.g4	22	25.1	65.4
rca1c.pk005.k3	24	48.8	67.1
ses8w.pk0020.b5	26	28.0	65.7
wl1n.pk0065.f2	28	29.8	67.9
rca1c.pk005.k3:fis	59	28.6	65.9

The amino acid sequence set forth in SEQ ID NO:24 is identical to amino acids 18 through 99 of the amino acid sequence set forth in SEQ ID NO:59.

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode a substantial portion of a corn and a wheat homoserine kinase, a portion and an entire rice homoserine kinase, and an entire soybean homoserine kinase.

EXAMPLE 6

Characterization of cDNA Clones Encoding Cysteine Synthase

The BLASTX search using the EST sequences from the clone listed in Table 12 revealed similarity of the polypeptides encoded by the cDNAs to cysteine synthase from *Citrullus lanatus* (DDJB Accession No. D28777, NCBI General Identifier No. 540497). Shown in Table 12 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones encoding the entire protein ("CGS"):

TABLE 12

BLAST Results for Sequences Encoding Polypeptides Homologous
to Cysteine γ Synthase

Clone	Status	BLAST pLog Score
		NCBI GI 540497 (<i>Citrullus lanatus</i>)
se3.05h06	CGS	182.64

Further sequencing and searching of the DuPont proprietary database allowed the identification of corn and rice clones encoding polypeptides with similarities to cysteine γ synthase. The BLAST search using the sequences from clones listed in Table 13 revealed similarity of the polypeptides encoded by the cDNAs to cysteine γ synthase from *Spinacia oleracea* (NCBI General Identifier No. 416869) and *Solanum tuberosum* (NCBI General Identifier No. 11131628). Shown in Table 13 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones encoding the entire protein ("CGS"):

TABLE 13

BLAST Results for Sequences Encoding Polypeptides Homologous
to Cysteine γ Synthase

Clone	Status	BLAST pLog Score	
		NCBI GI 416869 (<i>Spinacia oleracea</i>)	NCBI GI 11131628 (<i>Solanum tuberosum</i>)
Contig of: cco1n.pk083.j4 chp2.pk0016.b1 cpd1c.pk004.b20 cr1n.pk0083.c5 csi1.pk0003.g6 p0126.cnlcb49r rls6.pk0068.b7:fls	CGS	158.00	157.00
	CGS	161.00	163.00

Figure 5 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:31, 62, and 64 with the *Citrullus lanatus* sequence (NCBI General Identifier No. 540497; SEQ ID NO:32), *Spinacia oleracea* (NCBI General Identifier No. 416869; SEQ ID NO:65), and the *Solanum tuberosum* sequence (NCBI General Identifier No. 11131628; SEQ ID NO:66). The data in Table 14 presents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:31, 62, and 64 with the *Citrullus lanatus* sequence (NCBI General Identifier No. 540497; SEQ ID NO:32), *Spinacia oleracea* (NCBI General Identifier No. 416869; SEQ ID NO:65), and the *Solanum tuberosum* sequence (NCBI General Identifier No. 11131628; SEQ ID NO:66).

TABLE 14

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide
Sequences of cDNA Clones Encoding Polypeptides
Homologous to Cysteine γ Synthase

Clone	Amino acid SEQ ID NO.	Percent Identity to		
		NCBI GI 540497	NCBI GI 416869	NCBI GI 11131628
se3.05h06	31	87.1	72.3	76.9
Contig of:	62	73.8	71.3	69.7
cco1n.pk083.j4				
chp2.pk0016.b1				
cpd1c.pk004.b20				
crln.pk0083.c5				
csi1.pk0003.g6				
p0126.cnlcb49r				
rls6.pk0068.b7:fis	64	73.2	72.6	72.8

Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode entire corn, rice, and soybean cysteine γ synthases. These sequences represent the first corn, rice, and soybean sequences encoding cysteine γ synthase known to Applicant.

EXAMPLE 7

Characterization of cDNA Clones Encoding Cystathione β -Lyase

The BLASTX search using the EST sequences from clones listed in Table 15 revealed similarity of the polypeptides encoded by the cDNAs to cystathionine β -lyase from *Arabidopsis thaliana* (GenBank Accession No. L40511; NCBI General Identifier No. 1708993). Shown in Table 15 are the BLAST results for individual ESTs ("EST"), the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or the sequences of FISs encoding the entire protein ("CGS"):

TABLE 15

BLAST Results for Sequences Encoding Polypeptides Homologous
to Cystathione β -Lyase

Clone	Status	BLAST pLog Score 1708993 (<i>A. thaliana</i>)
cen1.pk0061.d4	FIS	50.41
rlr12.pk0026.g1	EST	39.00
sfl1.pk0012.c4	CGS	33.85
wrl.pk0091.g6	EST	52.52

The sequence of the entire cDNA insert in the clone wrl.pk0091.g6 was determined, RACE PCR was used to obtain the 5' portion of the rice cystathionine β -lyase, and further sequencing and searching of the DuPont proprietary database allowed the identification of other corn and wheat clones encoding cystathionine β -lyase. The BLASTX search using the EST sequences from clones listed in Table 16 revealed similarity of the polypeptides encoded by the cDNAs to cystathionine β -lyase from *Arabidopsis thaliana* (GenBank Accession No. L40511; NCBI General Identifier No. 1708993). Shown in Table 16 are the BLAST results for the sequences of the entire cDNA inserts comprising the indicated cDNA clones ("FIS"), or the sequences encoding the entire protein derived from contigs assembled from the sequences of more than two ESTs, the sequence of contigs assembled from the entire cDNA inserts comprising the indicated cDNA clones and 5' RACE PCR or an EST ("Contig*"):

TABLE 16

BLAST Results for Sequences Encoding Polypeptides Homologous
to Cystathione β -Lyase

Clone	Status	BLAST pLog Score 1708993
Contig of: cen1.pk0061.d4 p0005.cbmei71r p0014.ctuui39r p0109.cdadg47r p0125.czaay16r	Contig*	>180.00
5' RACE PCR+ rlr12.pk0026.g1: fis	Contig*	178.00
wrl.pk0091.g6: fis	FIS	177.00

Figure 6 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72 with the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 1708993; SEQ ID NO:41). The data in Table 17 presents a calculation of the

percent identity of the amino acid sequences set forth in SEQ ID NOs:34, 36, 38, 40, 68, 70, and 72 with the *Arabidopsis thaliana* sequence (NCBI General Identifier No. 1708993; SEQ ID NO:41).

TABLE 17

5 Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Cystathione β -Lyase

Clone	SEQ ID NO.	Percent Identity to 1708993 (<i>Arabidopsis thaliana</i>)
cen1.pk0061.d4	34	83.0
rlr12.pk0026.g1	36	76.0
sfl1.pk0012.c4	38	72.2
wr1.pk0091.g6	40	71.8
Contig of:	68	66.8
cen1.pk0061.d4		
p0005.cbmei71r		
p0014.ctuui39r		
p0109.cdadg47r		
p0125.czaay16r		
5' RACE PCR+	70	66.2
rlr12.pk0026.g1:fis		
wr1.pk0091.g6:fis	72	66.2

The amino acid sequence set forth in SEQ ID NO:34 is identical to amino acids 248 through 470 of the amino acid sequence set forth in SEQ ID NO:68. The amino acid sequence set forth in SEQ IDNO:36 is identical to amino acids 152 through 226 of the amino acid sequence set forth in SEQ ID NO:70. The amino acid sequence set forth in SEQ ID NO:40 is identical to amino acids 3 through 133 of the amino acid sequence set forth in SEQ ID NO:72.

15 Sequence alignments and percent identity calculations were performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for
20 pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. Sequence alignments and BLAST scores and probabilities indicate that the nucleic acid fragments comprising the instant cDNA clones encode an entire soybean cystathionine β -lyase, a substantial portion and an entire corn and
25 rice cystathionine β -lyases, a portion and asubstantial portion of a wheat cystathionine β -lyase.

EXAMPLE 8

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (Nco I or Sma I) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below.

Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes Nco I and Sma I and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb Nco I-Sma I fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sal I-Nco I promoter fragment of the maize 27 kD zein gene and a 0.96 kb Sma I-Sal I fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable

marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a KaptonTM flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a BiolisticTM PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the

tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

EXAMPLE 9

Expression of Chimeric Genes in Dicot Cells

5 A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about
10 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

15 The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

20 Soybean embryos may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and
25 placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night
30 schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein et al. (1987) *Nature* (London) 327:70-73, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used
35 for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al.

(1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 μ L of a 60 mg/mL 1 μ m gold particle suspension is added (in order): 5 μ L DNA (1 μ g/ μ L), 20 μ L spermidine (0.1 M), and 50 μ L CaCl_2 (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 μ L 70% ethanol and resuspended in 40 μ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five μ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 10

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with

additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

5 Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% low melting agarose gel. Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies, Madison, WI) according to the
10 manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs (NEB), Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and
15 deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect
20 to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately
25 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe
30 sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 11
Evaluating Compounds for Their Ability to Inhibit the Activity
of Plant Biosynthetic Enzymes

The polypeptides described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 10, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant polypeptides may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("His₆"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant polypeptides, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the instant polypeptides are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, the instant polypeptides may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a (His)₆ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β-mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the instant polypeptides disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. Examples of assays for many of these enzymes can be found in *Methods in Enzymology* Vol. V, (Colowick and Kaplan eds.) Academic Press, New York or *Methods in Enzymology* Vol. XVII, (Tabor and Tabor eds.) Academic Press, New York. Specific examples may be found in the following references, each of which is incorporated herein by reference: aspartic semialdehyde dehydrogenase may be assayed as described in Black et al. (1955) *J. Biol. Chem.* 213:39-50, or Cremer et al. (1988) *J. Gen. Microbiol.* 134:3221-3229; diaminopimelate decarboxylase may be assayed as described in Work (1962) in *Methods in Enzymology* Vol. V, (Colowick and Kaplan eds.) 864-870, Academic Press, New York or Cremer et al. (1988) *J. Gen. Microbiol.* 134:3221-3229; homoserine kinase may be assayed as described in Aarnes (1976) *Plant Sci. Lett.* 7:187-194; cysteine synthase may be assayed as described in Thompson et al. (1968) *Biochem. Biophys. Res. Commun.* 31: 281-286 or Bertagnolli et al. (1977) *Plant Physiol.* 60:115-121; and cystathionine β -lyase may be assayed as described in Giovanelli et al. (1971) *Biochim. Biophys. Acta* 227:654-670 or Droux et al. (1995) *Arch. Biochem Biophys.* 316:585-595.